Doing Bad to Look Good: Negative Consequences of Image Concerns on Pro-social Behavior

Ivan Soraperra, Anton Suvorov, Jeroen van de Ven, Marie Claire Villeval

GATE LYON SAINT-ÉTIENNE
UMR 5824

93, chemin des Mouilles
69130 Ecully - France

Maison de l'Université, Bâtiment B
10, rue Tréfilerie
42023 Saint-Etienne cedex 02 - France

http://www.gate.cnrs.fr
gate@gate.cnrs.fr

WP 1926 – September 2019

Doing Bad to Look Good: Negative Consequences of Image Concerns on Pro-social Behavior

Ivan Soraperra, Anton Suvorov, Jeroen van de Ven,
Marie Claire Villeval

Abstract:

Several studies show that social image concerns stimulate pro-social behavior. We study a setting in which there is uncertainty about which action is pro-social. Then, the quest for a better social image can potentially conflict with genuinely pro-social behavior. This conflict can induce \bad" behavior, where people lower both their own and others' material payoffs to preserve a good image. This setting is relevant for various types of credence goods. For example, recommending an inexpensive treatment reduces the expert's profits and may not satisfy the true needs of the client, but is generally good for the expert's image (as it signals the lack of greed). We test experimentally if people start to act bad in order to look good. We find that people care about their social image, but social image concerns alone do not induce them to act bad. That is, without future interactions, social image concerns do not lead to bad behavior. However, with future interactions, where building up a good image has instrumental value (reputational concerns), we do find evidence of bad behavior in the short run to secure higher earnings in the long run.

Keywords:

Social image, credence goods, prosocial behavior, reputation, experiment

JEL codes:
C92, D82, D91

Working Paper

Doing Bad to Look Good: Negative Consequences of Image Concerns on Pro-social Behavior*

Ivan Soraperra†, Anton Suvorov‡, Jeroen van de Ven§, and

Marie Claire Villeval¶

September 9, 2019

Abstract

Several studies show that social image concerns stimulate pro-social behavior. We study a setting in which there is uncertainty about which action is pro-social. Then, the quest for a better social image can potentially conflict with genuinely pro-social behavior. This conflict can induce "bad" behavior, where people lower both their own and others' material payoffs to preserve a good image. This setting is relevant for various types of credence goods. For example, recommending an inexpensive treatment reduces the expert's profits and may not satisfy the true needs of the client, but is generally good for the expert's image (as it signals the lack of greed). We test experimentally if people start to act bad in order to look good. We find that people care about their social image, but social image concerns alone do not induce them to act bad. That is, without future interactions, social image concerns do not lead to bad behavior. However, with future interactions, where building up a good image has instrumental value (reputational concerns), we do find evidence of bad behavior in the short run to secure higher earnings in the long run.

Acknowledgements: We are very grateful to Arthur Schram, Jean Tirole, and two anonymous referees for very helpful comments and suggestions. Financial support from the Research Priority Area Behavioral Economics at the University of Amsterdam (IS, JvV) and from the Van Gogh PHC program (IS, JvV, MCV) is gratefully acknowledged. A. Suvorov gratefully acknowledges financial support of FWF-RSF grant 18-48-05007 and thanks Toulouse School of Economics for hospitality. For MCV, this research has been conducted in the framework of the LABEX CORTEX (ANR-11-LABX-0042) of Universite de Lyon, within the program Investissements Avenir (ANR-11-IDEX-007) operated by the French National Research Agency (ANR).

†CREED, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands. E-mail: i.soraperra@uva.nl.

‡National Research University Higher School of Economics, Faculty of Economic Sciences, Pokrovsky bd., 11, Suite S1039; 109028 Moscow Russia. E-mail: asuvorov@hse.ru

§Tinbergen Institute and Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands. E-mail: j.vandeven@uva.nl.

¶Univ Lyon, CNRS, GATE UMR5824, 93 Chemin des Mouilles, F-69130, France; IZA, Bonn, Germany. E-mail: villeval@gate.cnrs.fr

1 Introduction

Although many people engage in pro-social behavior, they often do this out of image concerns rather than pure altruism. This is witnessed by studies showing that pro-social behavior increases when kind actions are made public. For instance, Andreoni and Petrie (2004) find that contribution levels in public good games increase when they are made public and the identity of the contributor is revealed. Ariely *et al.* (2009) find that people work harder when part of the proceeds go to charity and their efforts are publicly observable.

It is tempting to think that social image concerns will always result in more pro-social behavior. This, however, is not the case. Social image and pro-social behavior are not always aligned. This misalignment may happen when there is uncertainty about which action is pro-social and about the intentions of the decision-maker.[1] Ely and Valimaki (2003) illustrate this with the example of a car mechanic, a prime example in the literature on credence goods (*e.g.*, Dulleck and Kerschbamer, 2006). Clients may think that many car mechanics are opportunistic and will always perform unnecessary repairs. A mechanic that has good intentions and cares about his reputation may then want to avoid expensive repairs to signal that he is not a greedy opportunist. In more extreme cases, he may do this even when a more substantial repair is needed, thereby both hurting the client and reducing his own profits. Similar examples can be given for other credence goods, and related concerns exist in the realm of political advice (see Morris, 2001).

In this study, we examine if social image concerns can induce decision-makers to take actions that lower the payoffs for everyone in the context of credence goods provision. Such seemingly[2] Pareto-damaging behavior can be the result of social image concerns if there is some uncertainty about the appropriateness of a certain action and if that action is *usually* beneficial to the affected person. This behavior is similar to lying downward to avoid looking suspicious in the eyes of others.

We first present a simple model of a buyer-seller interaction that captures the main idea. Besides standard profit-seeking motives, the seller has some image concerns and a certain degree of professional pride or intrinsic motivation to do the right thing. These assumptions are in the spirit of Benabou and Tirole (2006) and Janssen and Mendys-Kamphorst (2004) who consider incentives for pro-social behavior. An important novelty in our model is that there is uncertainty about which action is pro-social.

[1] In the model of Benabou and Tirole (2006), "image rewards", while very effective in some situations, may not be practical in others; also, excessive signaling is possible even when it is common knowledge what action is "prosocial". DellaVigna et al.(20120) show that social pressure and image concerns may lead to substantial welfare losses in the context of charitable giving. We focus here on a different effect, when uncertainty about what action is prosocial may lead to antisocial, Pareto-damaging actions.

[2] "Seemingly", because it is defined in a narrow sense here, only taking into account immediate monetary payoffs. It it not Pareto-inefficient behavior if social image concerns and future rewards are taken into account.

The seller provides a credence good or service, and she is privately informed about whether the buyer needs a major or a minor repair. Providing a major repair is always beneficial for the seller, while the buyer only benefits from a repair that corresponds to her needs. Sellers are heterogeneous with respect to the strength of image concerns and professional pride. We show that if image concerns are moderate, the only possible kind of opportunism is the well-known over-treatment: sellers that lack intrinsic motivation provide a major repair when only a minor repair is needed. In contrast, when image concerns are strong in the population, sellers that are particularly image-sensitive may engage in image-driven under-treatment (providing a minor repair when a major one is needed). Such under-treatment lowers the immediate monetary payoffs for both parties, and in that sense is Pareto-damaging.

We then set up a laboratory experiment in which informed sellers performed a hypothetical repair for buyers. Sellers could choose between performing a major or a minor repair. For buyers, there was uncertainty about whether a major or minor repair was needed, but they knew that typically only a minor repair was needed. As is standard for credence goods, buyers cannot verify *ex post* whether a major or a minor repair was necessary. Performing a major repair was always beneficial for the sellers' monetary payoffs, independent of whether a minor or major repair was needed. The buyers' monetary payoffs were always higher if they got the repair that was actually needed. We varied whether sellers remained anonymous to the buyers and whether buyers could take some action after learning what kind of repair was performed. Our prediction was that performing a major repair would result in a worse social image, since that was the repair that was rarely needed but was more lucrative for the seller. Consequently, if sellers knew that a major repair was needed, they might nevertheless prefer to perform a minor repair, especially when their identity was revealed and even more when buyers could take some reciprocating action afterwards.

Our results show that participants in our experiment do indeed care about their social image, and that this can induce them to behave "badly", *i.e.*, to take the Pareto-damaging action. On the other hand, we only find this when building up a positive social image has an instrumental, material value that arises due to a possible reciprocal action by the buyer. When the choices of sellers are made public (but without verifiability of the buyers' true needs at any moment), and when buyers are given the opportunity to express their disapproval for the sellers' choices, sellers refrain from Pareto-damaging actions. However, when in a second stage buyers can allocate money between themselves and the seller, and thus reward sellers for good behavior, we find that sellers take the Pareto-damaging action 34% of the time (compared to 4% when buyers and sellers could also identify each other but without the possibility for the buyers to reward the seller).

Several papers relate to our work. One relevant strand of literature involves empirical work on credence goods. That body of research typically focuses on over-treatment

and over-charging (*e.g.*, Dulleck *et al.*, 2011; Balafoutas *et al.*, 2013; Beck *et al.*, 2014), and much less on under-treatment (see Schneider, 2012). The advantage of our experimental setting is that we can manipulate the value of reputation, and directly measure the type of the seller.

A closely related paper is the experimental study by Grosskopf and Sarin (2010). They also test for harmful effects of reputation in an experimental setting. Their setup closely resembles the model of "bad reputation" developed in Ely and Valimaki (2003). The main difference with our study is that they focus on a repeated game, in which long-run players interact with short run players and in which reputation building is purely instrumental. Grosskopf and Sarin (2010) do not find significant positive or negative effects of reputation. We also have a treatment in which reputation is instrumental, but add to this some treatments in which the value of reputation is purely intrinsic. We capture reputation not by means of a repeated game but in terms of social image. Our setup is also less complex, making it easier for participants to understand that bad behavior can help to build a good image, and does not involve some computerized players.

Ariely *et al.* (2009) also focus on the relationship between social image concerns and pro-social behavior. In line with the theoretical work by Benabou and Tirole (2006), Ariely *et al.* (2009) show that extrinsic motivation can crowd out image motivation. In a public setting, monetary rewards for pro-social behavior become an ineffective tool to stimulate donations. In their setting, monetary rewards can distort the relationship between pro-social behavior and building up a good image, but there is no scope to build up a good image by doing bad.

Another related strand of the literature focuses on communication in expert markets. Morris (2001) theoretically analyzes situations in which experts give advice to decision-makers. The expert can be good (her preferences are aligned with the decision-maker) or bad (she is opportunistic). After the first period of interaction, the decision-maker updates his beliefs about the expert's type. Morris shows that in the informative equilibrium, a good expert may lie (give the "politically correct" advice) to improve her future reputation. Chung and Harbaugh (2017) develop a related model and test this phenomenon experimentally. They find that when there is uncertainty about whether the expert is biased, even an unbiased expert will lie. They examine this in a setting where lying has instrumental value, rather than intrinsic.

Finally,[3] several recent studies discuss a related phenomenon to "doing bad to look good" in the context of lying. Those studies discuss the possibility of lying downward

[3]There is a also growing literature on the effects of social image concerns on economic behavior besides the references given above. See, for instance, a an excellent recent survey by Bursztyn and Jenssen (2017) and references therein. A few other recent papers are Karing (2018) and Karing and Naguib (2018) that demonstrate in field experiments that image concerns can be used to stimulate prosocial health-related behavior. Maskin and Tirole (2019) model a politician's tradeoff between overspending ("oversignaling") to pander to her support groups and being fiscally conservative ("undersignaling").

for subjects who suffer from perceived cheating aversion when they report the highest outcome of a random device. This type of behavior is driven by the willingness to appear honest when reporting the truth may look suspicious and people care about their reputation. Such under-reporting is ruled out in the model of Gneezy *et al.* (2018) and they do not find evidence of it in their set of experiments. Under-reporting is possible in the framework of Abeler *et al.* (2019) and Dufwenberg and Dufwenberg (2018). A difference with our setup is that those studies consider individual decision-making settings, without any social interaction and usually in contexts where the experimenter does not know the truth. This precludes the possibility of Pareto-damaging behavior. In a study involving asymmetric information between project managers and investors in financial markets and where some types of lies can be detected, Tergiman and Villeval (2019) find very little evidence of downward lying but without Pareto-damaging consequences.[4]

The paper is organized as follows. Section 2 presents our theoretical model. Section 3 introduces our experimental design and procedures. Section 4 reports our results and section 5 discusses these results and concludes.

2 Model

We model an interaction between a buyer and a seller. The seller has some degree of professional pride (or intrinsic motivation) and image motivation on top of standard pecuniary motives. This model is largely based on the approach of Benabou and Tirole (2006) who model pro-social behavior in the presence of image concerns and intrinsic motivation. An important novelty of our model is that we introduce uncertainty and private information regarding which action is actually "pro-social".

The buyer needs a repair of type $s \in \{0, 1\}$; where $s = 1$ corresponds to a major repair and $s = 0$ to a minor repair. It is common knowledge that a minor repair is needed with probability q, and a major one with probability $1 - q$. A minor repair is needed more often: $q > \frac{1}{2}$. The buyer does not know which kind of repair is needed, but the seller does. The seller chooses which repair to provide, $r \in \{0, 1\}$, where $r = 1$ corresponds to the major repair and $r = 0$ to the minor repair. The buyer always accepts the seller's offer (her outside option is 0).

The buyer's utility is given by:

$$U_B = a(1 - (r - s)^2),$$

where $a > 0$ reflects the buyer's gain from a correct repair. The incorrect repair has no value to the buyer. That there are only two possible values for the buyer (0 or a)

[4]The other studies that found some indirect evidence of downward lying (but in a non-strategic setting) are Utikal and Fischbacher (2013) with a sample of nuns, and Abeler *et al* (2014) in their telephone treatment.

is for simplicity only and is not important for the main insights of the model.[5]

The seller's utility is given by:

$$U_S = br + \theta(1 - (r - s)^2) + \alpha\eta\hat{\theta}_r.$$

The seller's monetary gain from a major repair is given by $b > 0$. The seller's monetary gain from performing a minor repair is normalized to zero. The parameter θ is a proxy for some combination of the seller's intrinsic motivation to provide the right kind of repair (his professional pride) and his altruism towards the buyer. The parameter α describes the common weight of the image concerns and η reflects the idiosyncratic strength of image concerns of a particular seller. Finally, $\hat{\theta}_r$ stands for the buyer's expectation of the seller's intrinsic motivation conditional on obtaining repair r.

Parameters a, b and α are common knowledge. In contrast, the true values of θ and η are known only to the seller, but it is common knowledge that θ and η are independent and uniformly distributed on $[0, 1]$. We assume that $b < 1$, so that the seller with the highest degree of intrinsic motivation and no image concerns ($\theta = 1, \eta = 0$) would always provide the right kind of repair.

The truthtelling (incentive compatibility) constraints for the seller are

$$\theta + \alpha\eta\hat{\theta}_0 \geq b + \alpha\eta\hat{\theta}_1 \tag{1}$$

and

$$b + \theta + \alpha\eta\hat{\theta}_1 \geq \alpha\eta\hat{\theta}_0 \tag{2}$$

for the cases when the minor and the major repair are needed respectively.

We first show that a minor repair results in a better social image for the seller.

Lemma 1. *In any perfect Bayesian equilibrium (PBE) the seller gets a strictly better social image when he provides a minor repair:* $\hat{\theta}_0 > \hat{\theta}_1$.

Proof. Assume that $\hat{\theta}_0 = \hat{\theta}_1$ in equilibrium. Then, (1) and (2) imply that sellers with $\theta \geq b$ provide the type of repair that the buyer needs and sellers with $\theta < b$ always provide a major repair. But then $\hat{\theta}_0 > \hat{\theta}_1$, contradicting our assumption.

If $\hat{\theta}_1 > \hat{\theta}_0$ in equilibrium, then the incentive constraints imply that the types of seller with $\eta \leq \frac{\theta - b}{\alpha(\hat{\theta}_1 - \hat{\theta}_0)}$ (which is a set with a positive mass under our assumptions) provide the required type of repair, and the other types of seller always choose a major repair. Then, using our assumption that θ and η are independent it can be easily checked that this would imply $\hat{\theta}_1 < \hat{\theta}_0$ in contradiction to the assumption. This is

[5] We also abstract away from pricing issues. In a more general setting, there could be different values for each combination of the state and repair, depending on the benefits of repairs and the prices charged. For instance, getting a major repair has sometimes value to the buyer even if only a minor repair is needed, since the major repair will also fix the minor problem. This could be (partially) offset by the higher price of a more expensive repair.

true for arbitrary absolutely continuous distribution functions; assumption about the uniform distribution is not needed for this result. □

The result of Lemma 1 is quite intuitive: given that a major repair is associated with a larger profit margin for the seller, obtaining this type of service casts doubts on the seller's motives.[6]

For each value of professional pride θ, denote by $\underline{\eta}(\theta)$ the value of image motivation that makes the seller indifferent between recommending a major and a minor repair when he knows that a minor repair is needed given the equilibrium beliefs. From (1), we get:

$$\underline{\eta}(\theta) = \frac{b - \theta}{\alpha(\hat{\theta}_0 - \hat{\theta}_1)}. \tag{3}$$

Similarly, denote by $\overline{\eta}(\theta)$ the value of image motivation that makes the seller indifferent between the two types of repair when he knows that a major repair is needed. From (2), we get:

$$\overline{\eta}(\theta) = \frac{b + \theta}{\alpha(\hat{\theta}_0 - \hat{\theta}_1)}. \tag{4}$$

Note that $\underline{\eta}(\theta)$ and $\overline{\eta}(\theta)$ may not belong to the interval $[0, 1]$. In that case, none of the types with a given θ is indifferent between the two types of repair.

Proposition 1. *There is always a PBE of the game. There exists a threshold value $\tilde{\alpha}$ such that:*

1. *If the weight of image concerns is weak enough ($\alpha < \tilde{\alpha}$), sellers with $\eta < \underline{\eta}(\theta)$ always offer a major repair regardless of the buyer's needs (area S_1 in Figure 1) and sellers with $\eta > \underline{\eta}(\theta)$) offer the kind of repair according to the buyer's true needs (area S_T in Figure 1).*

2. *If the weight of image concerns is strong enough ($\alpha \geq \tilde{\alpha}$), sellers with $\eta < \underline{\eta}(\theta)$ always offer a major repair regardless of the buyer's needs (area S_1 in Figures 2 and 3), sellers with $\underline{\eta}(\theta) \leq \eta \leq \overline{\eta}(\theta)$ offer the kind of repair according to the buyer's true needs (area S_T in Figures 2 and 3), and sellers with $\eta > \overline{\eta}(\theta)$ always offer a minor repair regardless of the buyer's needs (area S_0 in Figures 2 and 3).*

Proof. See the Appendix. □

Figure 1 illustrates the equilibrium seller's behavior when image concerns have a low weight in the seller's utility function. In this case the seller either gives an honest

[6] As intuitive as it is, this result does not generalize to the case when θ and η are not independent. To give an extreme example, assume that the distribution of types is discrete: there are two equally likely types of seller: $(b + \varepsilon, 0)$ and $(1, 1)$. Then, for α large enough there exists a PBE in which $(b + \varepsilon, 0)$-type does the required type of service, $(1, 1)$-type always makes a major repair and, as a result, $\hat{\theta}_1 > \hat{\theta}_0$. However, the truthful equilibrium also exists in this example.

advice to the buyer (if his intrinsic motivation is strong enough), or opportunistically suggests a major repair when only a minor one is needed.

Figure 1: Equilibrium with a low weight on image concerns ($\alpha < \widetilde{\alpha}$).

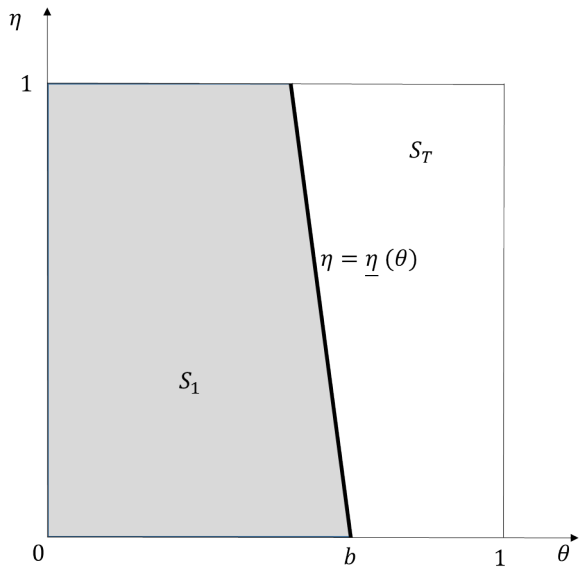


Figure 2: Equilibrium with a high weight on image concerns ($\widetilde{\alpha} < \alpha < \bar{\alpha}$).

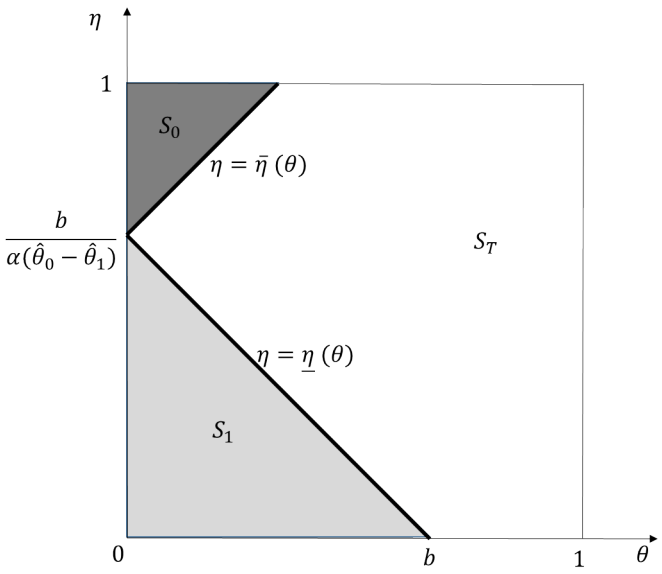
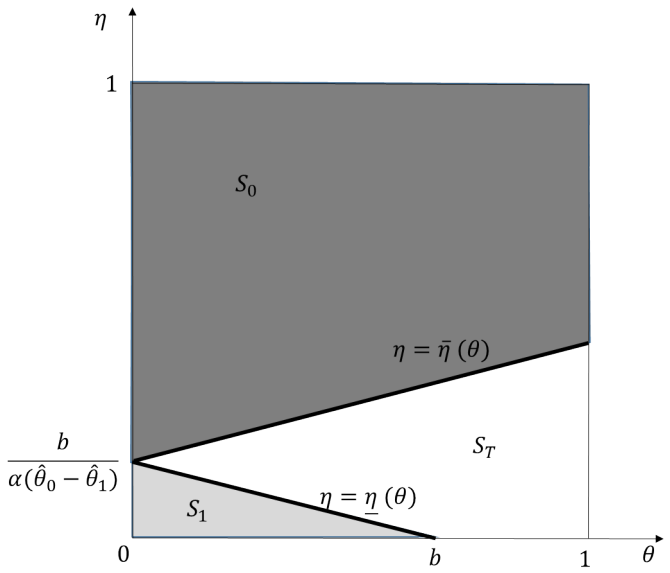


Figure 3: Equilibrium with a high weight on image concerns ($\alpha > \bar{\alpha}$).



When image concerns have a large impact on the seller's behavior, a conflict between image concerns and intrinsic motivation to do the right kind of repair becomes apparent and a new kind of opportunism emerges: the sellers with strong image concerns and low intrinsic motivation provide a minor repair even when the buyer needs a major one. These two cases are represented in Figures 2 and 3. The difference between the two cases is whether $\overline{\eta}(\theta)$ line crosses the upper or the right border of the unit square.

When the equilibrium corresponds to Figure 1 or Figure 2, for any level of image concerns η there exists a threshold level $\overline{\theta}(\eta) < 1$ such that the seller provides an adequate repair if $\theta > \overline{\theta}(\eta)$. In contrast, when Figure 3 describes equilibrium behavior, all sellers with high image concerns behave opportunistically regardless of their intrinsic motivation.

Thus far we have ignored any reciprocal actions by the buyer. In Appendix B, we extend our model to allow for reciprocity, assuming that buyers reward actions that they perceive as honest. This extended model better corresponds to the REWARD treatment in our experiment described below. We show that the introduction of a possibility to get a reward from the buyer affects the seller's behavior in a way similar to the stronger image concerns we modeled so far.

To summarize, the model gives the following testable predictions. First, a minor repair should result in a better social image than a major repair. Second, when image concerns are sufficiently strong, a new kind of opportunism appears: sellers with strong image concerns but low intrinsic motivation will under-treat the buyers, performing minor repairs when major repairs are needed.

3 Experimental Design and Procedures

Participants were divided into two roles that they kept throughout the experiment: sellers and buyers. Buyers were in the possession of a hypothetical product that needed a repair. The needed repair was either minor or major. Buyers were never informed whether a minor or a major repair was needed. They were told that this was determined by a random draw, and that a minor repair was needed 5/6 times and a major repair was needed 1/6 times.

Sellers were informed on which repair was needed, and had to choose between performing a minor or major repair. The payoffs are summarized in Table 1. Compared to a minor repair, a major repair always increased the seller's payoff by 1 (from 6 to 7 points), independent of whether a minor or major repair was needed. We kept this difference deliberately small, so that a major repair would be regarded as opportunistic behavior when a minor repair was needed. The buyer's payoff was 10 points if (s)he got the repair that was needed, and 5 if (s)he got the repair that was not needed. We used only two possible payoffs to keep the game simple to understand and to make it harder for the buyer to infer the outcome of the die roll from his or her earnings.

Table 1: Payoffs

State of the world	Seller's choice	
	Minor repair	Major repair
Minor repair needed (prob. = 5/6)	(6,10)	(7, 5)
Major repair needed (prob. = 1/6)	(6, 5)	(7,10)

Notes: Payoffs are denoted as (seller, buyer). The state of the world is determined by the roll of a die.

We implemented four treatments of this game, which varied in the anonymity between the players and in the buyer's options (passive or active), as detailed in Table 2.

In treatment "PRIVATE", the buyer and seller remained anonymous and the buyer was passive. This is essentially a dictator game in which the seller decides between allocations.

In treatment "SOCIAL WEAK", the buyer and seller were sitting next to each other in private cubicles. The seller was asked to hand over a card with his or her decision (but not specifying the outcome of the die roll). Each participant was also asked to indicate whether they were friends, acquaintances, or strangers. This ensured that participants were forced to look at each other and that the seller's identity was revealed to the buyer. This was meant to induce social image concerns (*i.e.*, to increase

the value of α of the model). All decisions were made in private. The buyer remained passive and had no action to take.

Treatment "SOCIAL STRONG" was similar to "SOCIAL WEAK" with two additional changes that were meant to further increase social image concerns. First, the buyer now rated the seller's decision on a scale from 0 to 10, and this was shown to the seller. The rating did not have any monetary consequences for the seller and therefore had no instrumental value, but possibly had an intrinsic value and forced some more interaction between the buyer and seller. Furthermore, at the end of all (five) rounds, each seller in SOCIAL STRONG was asked to stand up one by one and we publicly announced the number of times that they chose a major repair.

Treatment "REWARD" was similar to "SOCIAL WEAK", except that after the main game the buyer could divide 15 points between him- or herself and the seller, and these points were later converted to money. This gave buyers the opportunity to reward sellers for their decision. The buyers' decisions were not revealed to the seller.

Subjects were told that only one, randomly selected round would count for payment. We also told them that we would not reveal which round was selected. We did this so that buyers could not infer the seller's decision from his or her earnings, and to minimize the risk that they would redistribute earnings afterwards.

Table 2: Overview of treatments

Treatment	Anonymity	Public announcement of seller's decisions	Buyer's task
PRIVATE	Yes	No	Passive
SOCIAL WEAK	No	No	Passive
SOCIAL STRONG	No	Yes	Rate seller
REWARD	No	No	Allocate money

We conjectured that sellers would care about their image, either for instrumental reasons (in the REWARD treatment) or for intrinsic reasons (in the SOCIAL treatments). We also conjectured that the image concerns would be stronger in SOCIAL STRONG than in SOCIAL WEAK. Based on these conjectures, and the model's predictions, we formulate several hypotheses.

An immediate implication of Lemma 1 is that performing a minor repair ensures a better image than performing a major repair. Accordingly, we expect buyers to show more approval and allocate more money to sellers who perform a minor repair:

Hypothesis 1: Compared to sellers who perform a major repair, sellers who perform a minor repair get higher approval rates (in SOCIAL STRONG) and more money is

allocated to them (in REWARD).

The model also makes predictions about the fraction of pro-social and Pareto-damaging choices. If Figure 1 is relevant, we should not observe Pareto-damaging choices. One can show that in this case the fraction of pro-social choices increases in the strength of image concerns (as captured by a higher value of α or, for the extended model presented in Appendix B, of α or γ). If Figure 2 or 3 is relevant, then there will be Pareto-damaging choices. While for this case we have not shown that the fraction of pro-social and Pareto-damaging choices increases in the strength of image concerns[7], this case only becomes relevant for sufficiently strong image concerns in the first place. We therefore formulate the next two hypotheses:

Hypothesis 2: When a minor repair is needed, more sellers make pro-social choices in the REWARD condition and the SOCIAL conditions compared to the PRIVATE condition, and the effect is larger in SOCIAL STRONG than in SOCIAL WEAK.

Hypothesis 3: When a major repair is needed, more sellers make Pareto-damaging choices in the REWARD condition and the SOCIAL conditions compared to the PRIVATE condition, and the effect is larger in SOCIAL STRONG than in SOCIAL WEAK.

Procedures. Sessions were run in 2017 and 2019 as a classroom experiment in the CREED lab in Amsterdam. It was part of a course in experimental economics at the University of Amsterdam. In total, 176 subjects (51 percent female) participated in eight sessions. Each session had between 18 and 28 participants.

Subjects kept their role throughout the experiment and were re-matched after every round (total stranger design). We used a within-subject design. In 2017, each subject played 5 rounds in SOCIAL WEAK, 5 rounds in REWARD, and one round in PRIVATE.[8] The order of the treatments SOCIAL WEAK and REWARD was varied between sessions. The treatment PRIVATE always followed the others. In 2019, subjects played 3 rounds in SOCIAL WEAK and then 5 rounds in SOCIAL STRONG, followed by one round in PRIVATE. We did not vary the order of treatments in those sessions, because the public announcement of sellers' decisions in SOCIAL STRONG may otherwise affect sellers' behavior in SOCIAL WEAK afterwards.

We used the strategy method to elicit sellers' choices. This way we have information about their preferences for both contingencies (minor or major repair needed). After they indicated their choices for each contingency, we let sellers roll a die in front of

[7]The algebra is cumbersome, but numerical simulations we have done show a generally positive impact of increased image concerns on the likelihood of both Pareto-damaging and pro-social choices.

[8]In one session we could not complete the last 3 rounds of SOCIAL WEAK due to time constraints.

the experimenter to determine whether a minor or major repair was needed. We then handed them a slip of paper with the relevant seller's decision printed on it. Sellers handed over this slip of paper to the buyer.

As already mentioned, in each round (except for the PRIVATE treatment), subjects were asked to indicate if the paired participant was a stranger, acquaintance, or friend. The main purpose of this was to force subjects to look at each other, to create social image concerns among sellers. The reported answers are highly reciprocal; roughly 82 percent of the time the buyer and seller gave the same answer and in less than one percent of the cases was there a mismatch where one player indicated to be strangers and the other to be friends.

At the end of each round, sellers remained seated while buyers rotated such that each seller never interacted more than once with the same buyer. They were not allowed to communicate with each other.

The experiment was run using paper and pencil. Instructions can be found in Appendix C. At the end of the experiment we asked for the subject's gender. Sessions lasted between 75 and 90 minutes. Subjects were paid for one round randomly selected at the end of the session. The conversion rate was 1 euro for 1 point if the question was selected for payment. Subjects were not told which round was selected. Average earnings were 7.30 euros. There was no show up fee.

4 Results

We first verify that sellers' choices are informative of the state of the world. This is indeed the case, as illustrated in Table 3, which displays the distribution of sellers' choices by treatment and condition. Conditional on getting a minor repair, the empirical likelihood that a minor repair is actually needed is above 90 percent in all treatments. Conditional on getting a major repair, the likelihood that a major repair is actually needed is substantially lower, varying between 30 and 45 percent. A minor repair is therefore a reliable signal about the kind of repair truly needed, while a major repair is much less so.

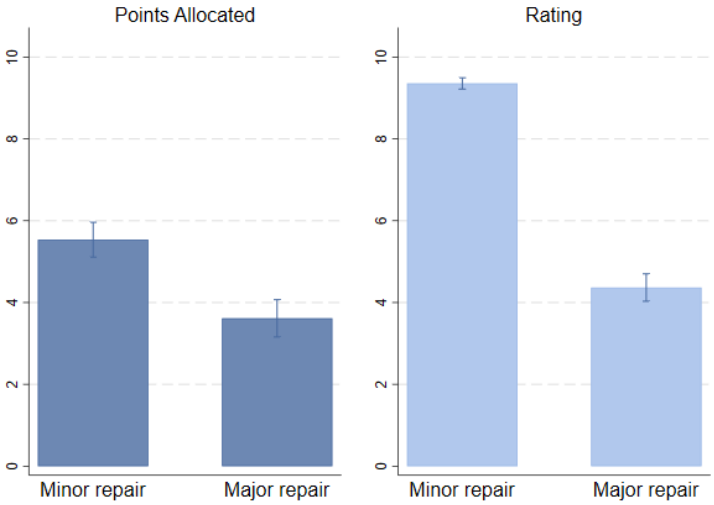
Table 3: Distribution of sellers' choices, by treatment and condition

Treatment	Seller's implemented choice	
	Minor repair	Major repair
Percentage of times minor needed		
SOCIAL WEAK	96	70
SOCIAL STRONG	97	60
REWARD	92	55

When buyers are active, their choices are consistent with the informational contents of the seller's implemented choice: Major repairs are generally not appreciated by buyers. Figure 4 shows how much they approve each type of repair and the amount of money that is allocated to sellers. In REWARD, buyers allocate on average 1.92 fewer points to the seller after getting a major repair compared to a minor repair. The difference is highly significant ($p < 0.001$, paired t-test)[9]. In SOCIAL STRONG, buyers rate sellers on average by 4.99 points less after getting a major instead of a minor repair ($p < 0.001$, paired t-test). These results support Hypothesis 1.

Result 1: Sellers build up a good image by performing a minor repair: buyers give higher approval ratings and allocate more money to sellers that offer a minor repair.

Figure 4: Left: Average number of points allocated by buyers in the REWARD treatment, by choice of the seller. Right: Mean rating by buyers in the SOCIAL STRONG treatment, by choice of the seller. Error bars indicate +/- 1 s.e.



What we have shown so far is that sellers can build up a more positive social image by performing a minor repair. We now turn to the question whether sellers' choices are driven by social image concerns. If so, we expect that stronger image concerns will increase pro-social choices when a minor repair is needed and an increase in Pareto-damaging choices when a major repair is needed.

When a minor repair is needed, sellers act pro-socially 37 percent of the time when

[9]Unless indicated otherwise, all statistical tests are based on the mean choice of an individual over all rounds as the unit of observation. Where applicable, we report results from paired tests, dropping buyers that always got a minor or always a major repair. Results from non-paired tests are very similar. Where the results differ, this is reported in the text.

their choices are anonymous, see Table 4. This increases to 83 percent in the REWARD condition ($p < 0.001$, paired t-test). Thus, sellers clearly care about their image when buyers have the option to reward them. They also care about their social image *per se*. Indeed, compared to the PRIVATE treatment, they act much more pro-socially in SOCIAL STRONG (an increase of 34 percentage points, $p < 0.001$, paired t-test $n = 46$). In SOCIAL WEAK, there is also an increase but only a modest one (12 percentage points, $p = 0.013$, paired t-test).[10]

Result 2: Compared to the PRIVATE treatment, sellers make more pro-social choices in SOCIAL WEAK and especially in SOCIAL STRONG and REWARD treatments.

Are subjects also willing to engage in Pareto-damaging behavior to preserve a good image? We find indeed that they do, but only when buyers can reward them. Pareto-damaging choices happen rarely when sellers remain anonymous (5 percent of the time in PRIVATE). In contrast, in the REWARD treatment we see a substantial increase of such behavior by 29 percentage points ($p < 0.001$, paired t-test). The SOCIAL WEAK treatment does not result in more Pareto-damaging choices ($p = 0.839$), and in the SOCIAL STRONG treatment the percentage of such choices increases by only 5 percentage points ($p = 0.071$, paired t-test[11]).

Result 3: Compared to the PRIVATE treatment, sellers make more Pareto-damaging choices when image has instrumental value (REWARD treatment), but not when image only has intrinsic value (SOCIAL WEAK and SOCIAL STRONG treatments).

Table 4: Fraction of pro-social and Pareto-damaging choices, by treatment.

Treatment	Pro-social choices	Pareto-damaging choices
PRIVATE	0.37	0.05
	(0.05)	(0.02)
SOCIAL WEAK	0.49	0.04
	(0.04)	(0.02)
SOCIAL STRONG	0.70	0.10
	(0.05)	(0.04)
REWARD	0.83	0.34
	(0.04)	(0.06)

Notes: s.e. in parentheses

[10] A non-paired t-test gives $p = 0.076$.
[11] A non-paired t-test gives $p = 0.198$.

Table 5 reports the results of a linear probability model with errors clustered at the seller and buyer level. The dependent variable is either the seller's choice of the pro-social option when a minor repair is needed (models (1) and (3)) or the seller's choice of the Pareto-damaging option when a major repair is needed (models (2) and (4)). The first two models include all pairs, while the last two models exclude from the sample pairs of friends. In these regressions, the PRIVATE treatment is the reference category.[12]

Table 5: Determinants of the sellers' probability to choose a minor repair.

	(1)	(2)	(3)	(4)
	Pro-social	Pareto-damaging	Pro-social	Pareto-damaging
Sample	All	All	Non-friends	Non-friends
SOCIAL WEAK (a)	0.119**	-0.001	0.069	0.002
	(0.048)	(0.024)	(0.049)	(0.025)
SOCIAL STRONG (b)	0.337***	0.054	0.317***	0.057
	(0.061)	(0.042)	(0.063)	(0.043)
REWARD	0.466***	0.295***	0.459***	0.317***
	(0.061)	(0.063)	(0.063)	(0.068)
Constant	0.368***	0.046**	0.368***	0.046**
	(0.052)	(0.023)	(0.052)	(0.023)
Wald Test (a)=(b) (p-value)	<0.001	0.135	<0.001	0.146
Observations	832	832	739	739
R-squared	0.112	0.127	0.126	0.138

Notes: Standard errors clustered at the seller and buyer level in parentheses. ***p<0.01, ** p<0.05, * p<0.1. Excluded treatment is PRIVATE.

Column (1) in Table 5 shows that all treatments significantly increase the likelihood of pro-social choices, with particularly large effects of REWARD and SOCIAL STRONG treatments. Column (2) indicates that only the REWARD treatment has a significant impact on the likelihood of Pareto-damaging choices, substantially increasing Pareto-damaging behavior. Columns (3) and (4) replicate these findings for the sample with friends excluded, yielding very comparable results.

Which sellers are most likely to engage in Pareto-damaging behavior? To answer this, we classify sellers according to their choices in the PRIVATE treatment. We classify a seller as "social" if he always chooses the repair that is needed in this treatment, and as "selfish" if he always chooses the major repair. 96 percent of all sellers falls into one of these categories, of which 36 percent are social and 60 percent selfish. We find that both types display Pareto-damaging behavior in REWARD treatment, although selfish types do this somewhat more often than social types; 38 percent and 25 percent

[12]Estimating alternatively a probit model delivers the same results.

choose the Pareto-damaging option, respectively, a difference that is not significant (p=0.331, two sample t-test).

In terms of consistency across rounds, we find that many subjects always take the same action in all rounds of a treatment. Still, a considerable fraction changes behavior across rounds. Most noticeable is that in treatment REWARD, about half of the sellers never engage in Pareto-damaging behavior. About 17 percent chooses the Pareto-damaging option in all rounds. The remaining sellers are roughly equally distributed. Figures 5 and 6 show the distributions for all treatments.

Figure 5: Histograms of the percentage of pro-social choices by each seller.

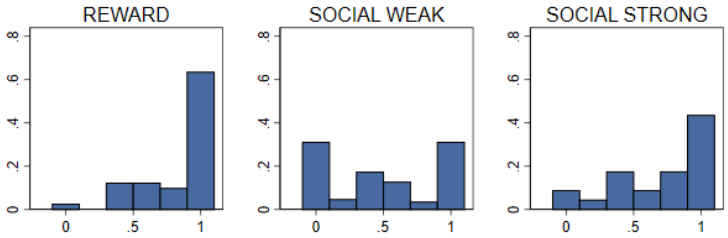
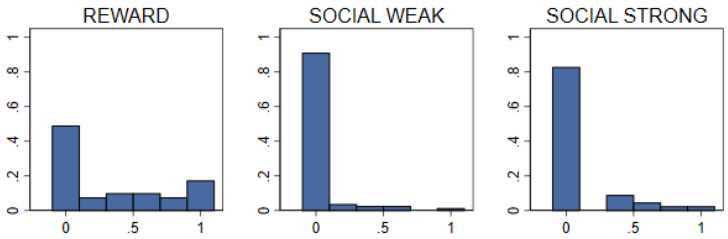
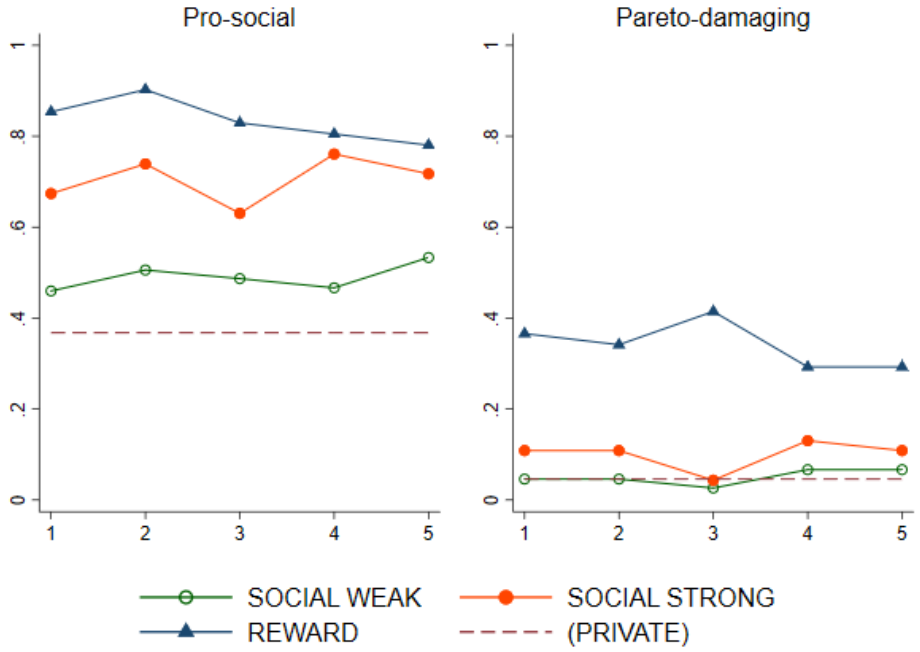


Figure 6: Histograms of the percentage of Pareto-damaging choices by each seller.



It is conceivable that subjects need some experience to understand how buyers perceive their choices. Learning might therefore take place, especially in the treatments in which sellers receive feedback from the buyers (REWARD and SOCIAL STRONG treatments). We find little evidence to support this. Figure 7 shows the fraction of pro-social and Pareto-damaging choices over rounds. Treatment differences are already visible in round 1, and there are no strong time trends.

Figure 7: Evolution of the fraction of pro-social and Pareto-damaging choices over time, by treatment. The dashed line represents the PRIVATE treatment (for which there is only a single round).



5 Concluding Remarks

In this paper, we provide a model and test experimentally whether social image concerns can play against pro-social actions in a context typical of credence goods, when informed sellers can decide to provide either a major or a minor repair depending on the uninformed buyer's needs. Our experimental results show that people indeed care about their social image, although image concerns alone do not induce them to sacrifice surplus. However, if the loss of image entails potential monetary consequences, we do find evidence of Pareto-damaging behavior. To preserve their social image when the latter is misaligned with pro-social behavior, some sellers act badly, providing a minor repair when a major one was needed to not appear greedy in the setting where buyers can allocate additional earnings after receiving their treatment. This behavior shares some common features with downward lying in cheating experiments where individuals under-report their outcome to not appear suspicious, but here in a social interaction context.

The question remains why we do not document Pareto-damaging behavior when social image concerns only has intrinsic value. Possibly, the social image concerns

that we induced were not strong enough even in the SOCIAL STRONG treatment. We cannot exclude this, but it does not seem to be in accordance with the fact that subjects clearly cared about it, as witnessed by the large increase in pro-social choices. That said, it is conceivable that the manipulation was not strong enough, and that the area for which Pareto-damaging behavior should be observed (*i.e.*, S_0 in Figures 2 and 3) is still too small to detect. Another possibility could be that there is some interaction between image concerns and other-regarding preferences, such that people care about their social image but are not willing to hurt others to preserve it.[13]

These results are important because they invite to remain cautious about the temptation to believe that social image concerns always favor pro-social actions. They are also important because they provide clear evidence of downward deception for strategic reasons, a behavior that so far has been very difficult to capture empirically. However, we acknowledge a number of limitations of this study. The most important one is that the experiment has been conducted as a classroom experiment, which may reinforce the concern for social image toward the experimenter. Note, however, that this particular setting may lead us to underestimate the importance of Pareto-damaging behavior if players refrained from this because of the experimenter scrutiny. Another concern might be *ex post* sharing of earnings. We minimized this risk by not revealing for which round they were paid, so that afterwards they could not identify their paired player, and by not letting them communicate verbally, so that they could not make any promises during the experiment. The fact that our results are robust to excluding pairs of participants that are friends suggests that these potential limitations were not severe. Also, note that a classroom experiment has an advantage in a context like ours. Since participants know each other, they are more likely to care about their social image.

Despite these limitations, we believe that our results highlight an intriguing phenomenon that should motivate further investigation of when and in which circumstances individuals may be willing to misbehave to look good in the eyes of others.

[13]We thank a reviewer for suggesting this.

References

Abeler, Johannes, Anke Becker, and Armin Falk (2014). Representative evidence on lying costs. *Journal of Public Economics*, 113: 96-104.

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond (2019). Preferences for truth-telling. *Econometrica*, 87(4): 1115-1153.

Andreoni, James, and Ragan Petrie (2004). Public goods experiments without confidentiality: a glimpse into fund-raising. *Journal of Public Economics*, 88(7-8), 1605-1623.

Ariely Dan, Anat Bracha, and Stephen Meier (2009). Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially. *American Economic Review* 99(1): 544-555.

Balafoutas, Loukas, A. Beck, Rudolph Kerschbamer, and Matthias Sutter (2013). What drives taxi drivers? A field experiment on fraud in a market for credence goods. *Review of Economic Studies*, 80: 876–891.

Beck, A., Rudolph Kerschbamer, J. Qiu, and Matthias Sutter (2014). Car mechanics in the lab - Investigating the behavior of real experts on experimental markets for credence goods. *Journal of Economic Behavior & Organization*, 108: 166–173.

Benabou Roland and Jean Tirole (2006). Incentives and prosocial behavior. *American Economic Review* 96(5): 1652-1678.

Bursztyn, Leonardo and Robert Jensen (2017). Social Image and Economic Behavior in the Field: Identifying, Understanding, and Shaping Social Pressure *Annual Reviews in Economics* 9: 131-53.

Chung, Wonsuk and Rick Harbaugh (2017). Biased Recommendations from Biased and Unbiased Experts. Mimeo, Kelley School of Business, Indiana University.

DellaVigna Stefano, John List and Ulrike Malmendier (2017). Testing for Altruism and Social Pressure in Charitable Giving. *Quarterly Journal of Economics* 127(1): 1-56.

Dufwenberg, Martin, and Martin A. Dufwenberg (2018). Lies in Disguise – A theoretical Analysis of Cheating. *Journal of Economic Theory,* 175: 248-264.

Dufwenberg, M., Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Eonomic Behavior*, 47(2), 268-298.

Dulleck, Uwe, and Rudolph Kerschbamer (2006). On doctors, mechanics, and computer specialists: The economics of credence goods. *Journal of Economic Literature*, 44: 5-42.

Dulleck, Uwe, Rudolph Kerschbamer and Matthias Sutter (2011). The economics of credence goods: An experiment on the role of liability, verifiability, reputation, and competition. *American Economic Review*, 101(2), 526-55.

Ellingsen, T., Johannesson, M. (2008). Pride and prejudice: The human side of incentive theory. *American Economic Review*, 98(3), 990-1008.

Ely, Jeffrey C., and Juuso Välimäki (2003). Bad Reputation. *The Quarterly Journal of Economics*, 785-814.

Falk, A., Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, 54(2), 293-315.

Fehr, E., Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114(3), 817-868.

Gneezy, Uri, Agne Kajackaite and Joel Sobel (2018). Lying Aversion and the Size of the Lie. *American Economic Review*, 108 (2): 419-453.

Grosskopf, Brit, and Rajiv Sarin (2010). Is reputation good or bad? An experiment. *American Economic Review*, 100: 2187-2204.

Janssen, Maarten and Ewa Mendys-Kamphorst (2004). The price of a price: on the crowding out and in of social norms. *Journal of Economic Behavior & Organization*, 55: 377-395.

Karing, Anne (2018). Social Signaling and Childhood Immunization: A Field Experiment in Sierra Leone. Mimeo.

Karing, Anne and Karim Naguib (2018). Social Signaling and Prosocial Behavior Experimental Evidence in Community Deworming in Kenya. Mimeo.

Maskin, Eric and Jean Tirole (2019). Pandering and Pork-Barrel Politics. *Journal of Public Economics*, 176, 79-93.

Morris, Stephen (2001). Political correctness. *Journal of Political Economy*, 109(2), 231-265.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, 1281-1302.

Schneider, Henry (2012). Agency problems and reputation in expert services: Evidence from auto repair. *The Journal of Industrial Economics*, 60(3), 406-433.

Tergiman, Chloe, and Marie Claire Villeval (2019). The Way People Lie in Markets: the Role of Reputation and Competition. Mimeo, GATE.

Utikal, Verena, and Urs Fischbacher (2013). Disadvantageous lies in individual decisions. *Journal of Economic Behavior & Organization*, 85: 108-111.

A Proofs of main propositions

Proof of Proposition 1. For each value of $\Delta\hat{\theta} = \hat{\theta}_0 - \hat{\theta}_1$, functions $\underline{\eta}(\theta)$ and $\overline{\eta}(\theta)$ defined in equations (3) and (4) determine the seller's behavior: the seller always provides a major repair if $\eta < \underline{\eta}(\theta)$, always provides a minor repair if $\eta > \overline{\eta}(\theta)$ and provides a repair that corresponds to the buyer's needs otherwise. Thus, for each value of $\Delta\hat{\theta}$ we can compute functions $\underline{\eta}(\theta)$ and $\overline{\eta}(\theta)$ and use Bayes rule to compute the updated values for $\hat{\theta}_0$ and $\hat{\theta}_1$. The equilibrium corresponds to a fixed point in the map $g : \Delta\hat{\theta} \to \hat{\theta}_0 - \hat{\theta}_1$ thus defined.

Note that for $\Delta\hat{\theta} < b/\alpha$, Figure 1 represents a potential equilibrium configuration (that is, area S_0 is absent). Then, we have[14]

$$\hat{\theta}_0 = E[\theta|S_T],$$

$$\hat{\theta}_1 = \frac{S_1 E[\theta|S_1] + (1-q)S_T E[\theta|S_T]}{S_1 + (1-q)S_T},$$

so that

$$\hat{\theta}_0 - \hat{\theta}_1 = \frac{S_1}{S_1 + (1-q)S_T}(E[\theta|S_T] - E[\theta|S_1]).$$

In particular, $\hat{\theta}_0 - \hat{\theta}_1 \to \frac{b}{2[1-q(1-b)]} > 0$ as $\Delta\hat{\theta} \to 0$.

On the other hand, when $\Delta\hat{\theta} = 1$ the potential equilibrium is represented by Figure 1 if $\alpha < b$, by Figure 2 if $b < \alpha < b+1$ and by Figure 3 if $\alpha > b+1$. In any case, both updated $\hat{\theta}_0$ and $\hat{\theta}_1$ are strictly between 0 and 1, so that $\hat{\theta}_0 - \hat{\theta}_1 < 1$. The map $h : \Delta\hat{\theta} \to \Delta\hat{\theta} - g(\Delta\hat{\theta})$ is obviously continuous on $[\varepsilon, 1]$ for any $\varepsilon \in (0,1)$. Moreover, the preceding analysis shows that it assumes a negative value at $\Delta\hat{\theta} = \varepsilon$ for any $\varepsilon \in (0,1)$ small enough and a positive value at $\Delta\hat{\theta} = 1$. Therefore, the map $g(\cdot)$ has a fixed point on that interval which corresponds to the PBE of the game.

Finally, denote by z the joint intersection of the $\underline{\eta}(\theta)$ and $\overline{\eta}(\theta)$ with the vertical axis; z is a decreasing function of α and $\Delta\hat{\theta}$. Since the area of S_1 is increasing in z, and the area of S_T is decreasing, $\frac{S_1}{S_1+(1-q)S_T}$ is a decreasing function of α and $\Delta\hat{\theta}$. One can also check that $E[\theta|S_T] - E[\theta|S_1]$ is an increasing function of z for $z \geq 1$ (when the equilibrium configuration corresponds to Figure 1). Hence, function $g(\cdot)$ is decreasing in $\Delta\hat{\theta}$ on the interval $(0, \frac{b}{\alpha}]$; moreover, it is decreasing in α for a fixed $\Delta\hat{\theta}$. Therefore, there exists $\tilde{\alpha}$ such that the equilibrium is like in Figure 1 (without Pareto-damaging opportunism, i.e., without S_0 region) for $\alpha < \tilde{\alpha}$ and it involves some Pareto-damaging opportunism (Figure 2 or Figure 3) for $\alpha > \tilde{\alpha}$.

□

[14] For brevity we write $E[\theta|S_j]$ instead of $E[\theta|(\theta,\eta) \in S_j]$ for $j \in \{0,1,T\}$. With a slight abuse of notation S_j denotes both a subset of parameters and the measure of this subset.

B Extended model with buyers' reciprocity

We extend our model to allow for the possibility that the buyer can take a reciprocal action to reward the seller ex post for behavior that he considers honest. This could be driven by some kind of social preferences (e.g., à la inequity aversion as in Fehr and Schmidt, 1999) or reciprocity (à la Rabin, 1993; Falk and Fischbacher, 2006; or Dufwenberg and Kirchsteiger, 2004). We do not model the buyer's preferences explicitly (this would take us beyond the scope of the paper), but augment instead the seller's payoff by introducing a potential reward from the buyer. Namely, there is now an additional term $\gamma\hat{\theta}$ with $\gamma > 0$ being some positive constant. We thus assume that the expected size of the buyer's reward increases linearly with her belief about the seller's honesty (in the spirit of Ellingsen and Johannesson, 2008, we believe that the buyer's willingness to reward the seller depends on her beliefs about how good/honest a person the seller is). We assume that the expected size of the buyer's reward is not too large compared to the material benefits of providing a major repair, i.e., $\gamma < b$.[15]

The difference with the original model is that the new term is not sensitive to η, the strength of the seller's image concerns. This complicates the analysis considerably and we will restrict ourselves to a few numerical examples, without the formal proofs. The examples suggest that the key insights remain intact.

Lemma 1 remains valid; the proof can be repeated verbatim. Figures 1-3 still depict potential equilibrium configurations, with the qualification that now

$$\underline{\eta}(\theta) = \frac{b - \gamma(\hat{\theta}_0 - \hat{\theta}_1) - \theta}{\alpha(\hat{\theta}_0 - \hat{\theta}_1)}; \tag{5}$$

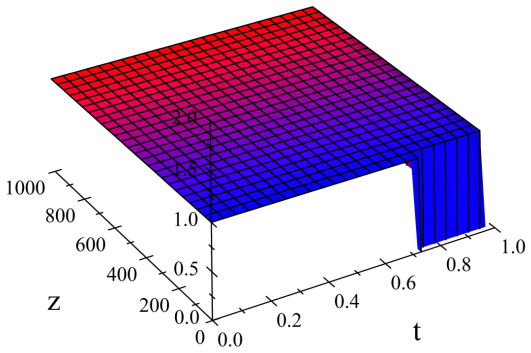
$$\overline{\eta}(\theta) = \frac{b - \gamma(\hat{\theta}_0 - \hat{\theta}_1) + \theta}{\alpha(\hat{\theta}_0 - \hat{\theta}_1)}; \tag{6}$$

the intersection of $\underline{\eta}(\theta)$ with the horizontal axis (denote it by t) is now $b - \gamma(\hat{\theta}_0 - \hat{\theta}_1)$ instead of b and the joint intersection of the $\underline{\eta}(\theta)$ and $\overline{\eta}(\theta)$ with the vertical axis (denote it by z) is $\frac{b-\gamma(\hat{\theta}_0-\hat{\theta}_1)}{\alpha(\hat{\theta}_0-\hat{\theta}_1)}$ instead of $\frac{b}{\alpha(\hat{\theta}_0-\hat{\theta}_1)}$. We cannot replicate directly the proof of Theorem 1. Before, with $\gamma = 0$, the effect of increase in the strength of image concerns, α, or the expected image benefit $\Delta\hat{\theta} = \hat{\theta}_0 - \hat{\theta}_1$ was an increase in z that lead to unambiguous fall in $g(\Delta\hat{\theta})$ whenever $\alpha\Delta\hat{\theta} \leq 1$, i.e. $z \geq 1$. Now an effect of increase in α is still an increase in z with t remaining constant; therefore, $g(\Delta\hat{\theta})$ falls in α. However, an increase in $\Delta\hat{\theta}$ or in the new parameter γ leads not only to a fall in z but also to a fall in t. While the first effect leads to a fall in $g(\Delta\hat{\theta})$, the second effect is the opposite. We cannot give a clear analytical proof that overall this leads to a fall in $g(\Delta\hat{\theta})$ which would allow to extend the proof of Theorem 1 to this new richer

[15] Without this assumption a new type of equilibrium could emerge, with the only type of opportunism occurring with a positive probability being Pareto-damaging behavior, i.e., the provision of a minor repair irrespective of the buyer's needs.

model. Instead, we take several specific values of q and verify numerically[16] that this is the case provided t is not too large. For instance, for $q = 5/6$ as in the example we use in the experiment the graph of $signum\left(\frac{d}{dt}g(\Delta\hat{\theta})\right)$ for values of α and γ such that $z > 1$ (i.e., Figure 1 is relevant) is given in Figure 8.

Figure 8:



The assumption that b is small enough (smaller than 0.7 for this particular example) is sufficient to ensure that an increase in any sort of image concerns, α or γ, leads to a fall in $g(\Delta\hat{\theta})$. All other numerical examples that we considered (with different values of q between 0.5 and 1) give qualitatively the same picture.

Thus, we have shown (partly, numerically) that for the moderate level of material interest b Theorem 1 can be extended: Pareto damaging behavior emerges if image concerns α or γ are strong enough.

[16]For this purpose we used the math engine of the ScientificWorkplace package.

C Instructions – PRIVATE treatment

Below are the instructions for sessions that started with the SOCIAL WEAK treatment. Instructions for the other sessions are available upon request.

Welcome

Welcome to this experiment. Please read these instructions very carefully. You are not allowed to communicate with other participants during the experiment.

You can earn money in this experiment. The amount depends on your own choices and the choices of other participants. Your earnings will be denoted in points.

This experiment has three parts, each with one or more rounds. At the end of the experiment, one of the rounds of one of the parts will be randomly selected for payment. Every point that you earned in that round is worth €1,00. Every part is equally likely to be selected for payment.

[SOCIAL WEAK] Description of the task for Part 1

There will be two types of players: Sellers and Buyers. Your role is indicated on your badge.

The Buyer has a hypothetical product that needs to be repaired. The repair that is needed is either "minor" or "major." Whether a minor or major repair is needed depends on the outcome of a die roll:

- If the outcome of the die roll is 1, 2, 3, 4, or 5, then only a minor repair is needed.
- If the outcome of the die roll is 6, then a major repair is needed.

The outcome of the die roll will only be visible to the Seller (and the experimenter), and not to the Buyer. This means that only the Seller knows whether a minor or major repair is needed.

The Seller's task is to perform a (hypothetical) repair. The Seller can choose to perform either a minor or a major repair. The payoffs are as follows:

- Choosing a major repair always gives the Seller one more point than a minor repair (7 instead of 6), independent of whether a minor or major repair is needed.
- The buyer earns five more points if the seller chooses to perform the repair that is needed. If the die roll is 1, 2, 3, 4, or 5, then only a minor repair is needed. In this case, choosing a minor repair gives 10 points and a major repair gives 5 points to the buyer. If the die roll is 6, then a major repair is needed. In this case, choosing a major repair gives 10 points and a minor repair gives 5 points to the buyer.

The Buyer will know whether the Seller chooses to perform a minor or major repair. However, since the Buyer does not observe the outcome of the die roll, (s)he cannot know for certain whether a minor or major repair is needed. All that the buyer knows is that there is a 1/6 chance that a major repair is needed, and a 5/6 chance that a minor repair is needed.

The earnings are summarized in the table below. The Buyer does not learn his or her payoffs until the end of the experiment, and we will not reveal which round was selected for payment, so he or she will not be able to tell from his or her earnings whether a minor or major repair gave more points in any specific round. The Seller's payoffs are in blue, and the Buyer's earnings are in orange.

Overview of Payoffs

	The outcome of the die roll is…	
	1,2,3,4,5 (needs minor repair)	6 (needs major repair)
The seller performs a minor repair	6 10	6 5
The seller performs a major repair	7 5	7 10

In each round, there are 3 steps.

Step 1: For each possible outcome of the die roll (1-5 or 6), the Seller indicates whether (s)he prefers to perform a minor or major repair.

Step 2: After the Seller made his or her choices, (s)he rolls a die. This determines which of the choices is relevant. The seller receives a card that indicates his or her choice (minor or major) for the outcome of the die roll. The card does not indicate the outcome of the die roll.

Step 3: The Seller hands over the card with his or her choice to the Buyer, who is sitting to the right of the Seller. *You are not allowed to say anything.*

On the answer sheet, we also ask you to indicate if the paired participant is a friend, acquaintance, or someone you don't know.

There will be 3 [5 in some sessions] rounds like this. After round 3, you will receive new instructions for Part 2. You will never meet the same Buyer or Seller again during the entire experiment. In every round, you will be matched with another participant.

Sellers always stay seated where they are now. Buyers will rotate after every round and move to the next Seller.

Brief Summary

- A die roll determines whether a minor or major repair is needed, with a 5/6 chance that a minor repair is needed
- The Seller observes the outcome of the die roll
- The Seller chooses to perform a minor or major repair
- The Seller earns one more point if (s)he chooses a major repair
- The Buyer earns five more points if the Seller chooses the repair that is needed
- The Buyer does observe the Seller's choice but not the outcome of the die roll

Please raise your hand if you have any questions.

[SOCIAL STRONG] Description of the task for Part 2

The instructions for part 2 are similar to those for part 1, except that in the following rounds, there is an additional step.

After observing the Seller's choice, the Buyer will rate the Seller's choice on a scale from 0 (very dissatisfied) to 10 (very satisfied).

Thus, in each round there are now 4 steps.

Steps 1 to 3: as before,

Step 4: The Buyer rates the Seller's choice and hands this over to the Seller.

There will be 5 rounds in this part.

At the end of the last round, we disclose the Seller's decisions to all the other participants. The experimenter will call out each Seller. The Seller has to stand up while the experimenter publicly announces how often the Seller chose a major repair and how often the Seller chose a minor repair. Only the Seller's decisions in this part will be made public, and only the decisions that were relevant to the Buyer (i.e., the decisions that followed after rolling the die).

Please raise your hand if you have any questions.

[REWARD] Description of the task for Part 2

The instructions for part 2 are similar to those for part 1, except that in the following rounds, there is an additional step. After observing the Seller's choice, the Buyer will have 15 points to divide between him- or herself and the paired Seller. The Buyer can choose to divide the 15 points as (s)he wishes. The Buyer will do this in private, and the Seller will not learn how (s)he divided the points unless this round is selected for payment at the end of the experiment.

Thus, in each round there are now 4 steps.

Steps 1 to 3: as before,

Step 4: The Buyer divides 15 points between her(him)self and the Seller.

There will also be 5 rounds in this part. If one of these rounds is selected for payment, it will be randomly determined which decision counts for your earnings: The seller's choice or the buyer's choice.

Please raise your hand if you have any questions.

[PRIVATE] Description of the task for Part 3

This is the last part. Please answer the following questions.

You are randomly paired with another **anonymous** participant from this group. If this part is selected for payment, it is randomly determined which of the two questions will count, and whether your decision counts or the decision of the paired participant.

1) Which option do you prefer?

- ☐ Option A: 6 points for you, 5 for the other
- ☐ Option B: 7 points for you, 10 for the other

2) Which option do you prefer?

- ☐ Option A: 7 points for you, 5 for the other
- ☐ Option B: 6 points for you, 10 for the other